

# Dynamic Clustering of Multivariate Panel Data

A quick overview

NBER-NSF Time Series Conference 2021

Igor Custodio João, André Lucas, Julia Schaumburg

Vrije Universiteit Amsterdam and Tinbergen Institute

Bernd Schwaab

European Central Bank, Financial Research

The views expressed in this presentation are those of the authors and they do not necessarily reflect the views or policies of the European Central Bank.

# Goal

- Exploit the panel structure to produce a flexible, time-varying clustering.
  - A Hidden Markov Model is used for the cluster transitions.
  - A mixture model with time-varying parameters is used for the observations.
- An application to bank data exemplifies the usefulness for regulatory supervision.

# The model in a few equations

Mixture model of the data:

$$\mathbf{y}_{it} = \boldsymbol{\mu}_{c_{it},t} + \boldsymbol{\varepsilon}_{it}, \quad \boldsymbol{\varepsilon}_{it} | c_{it} \sim t(\mathbf{0}, \boldsymbol{\Sigma}_{c_{it},t}, \nu_{c_{it}})$$

and time-varying mean and variances of the form:

$$\boldsymbol{\mu}_{j,t+1} = \boldsymbol{\mu}_{jt} + \mathbf{A}_1 \mathbf{S}_{\boldsymbol{\mu}_{jt},t} \cdot \nabla \boldsymbol{\mu}_{jt,t}$$

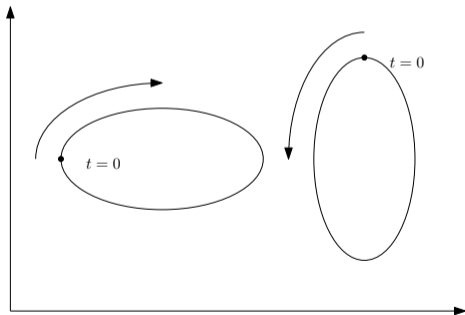
with  $c_{it}$  being cluster assignments following the HMM:

$$\mathbb{P}(c_{i,t+1} = k | c_{it} = j) = \frac{\exp(-\gamma \tilde{d}_{jk,t-1})}{\sum_{q=1}^J \exp(-\gamma \tilde{d}_{jq,t-1})}$$

where  $\tilde{d}_{jq,t}$  is a distance measure between cluster means  $\boldsymbol{\mu}_{jt}$  and  $\boldsymbol{\mu}_{kt}$ .

# Simulation

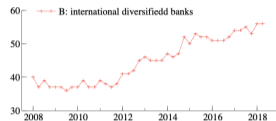
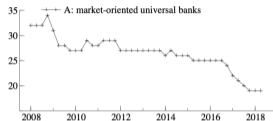
- We simulate elliptical trajectories for two clusters.
- We greatly outperform hierarchical clustering both on classification and tracking of the means.



dist.	$\sigma_2^2$	$\gamma$	Heterogeneous $A_1$				Hierarchical	
			$\hat{\gamma}$	%C	MSE	LL	%C	MSE
4	1	0.25	0.234	0.949	0.096	-898.8	0.729	2.161
4	1	0.50	0.470	0.963	0.090	-853.4	0.727	2.211
4	8	0.25	0.314	0.781	0.616	-1069.7	0.665	4.474
4	8	0.50	0.561	0.830	0.307	-1034.0	0.674	4.251
8	1	0.25	0.246	0.998	0.083	-875.5	0.883	0.796
8	1	0.50	0.493	0.999	0.083	-827.7	0.882	0.802
8	8	0.25	0.263	0.991	0.144	-1069.8	0.894	0.871
8	8	0.50	0.536	0.995	0.141	-1023.1	0.898	0.834

# Empirical application

- 12 bank-level accounting indicators from SNL Financial 2008Q1-2018Q2.
- 6 clusters, selected by well-known indices (silhouette and Davis-Bouldin).
- We find clusters with strong trends in their composition.
- They roughly align with the business models identified by the ECB's SSM.
- The trends can be explained by the profitability of each business model.



# Conclusion

- We introduce a novel model for studying time-varying group structures in multivariate panel data.
- It combines unsupervised learning (of the unknown groups) with the time-varying parameters models.
- It produces accurate results in simulations.
- The application to bank business models provides new insights into the dynamics of the sector in Europe.